REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
21-12-2015	Conference Proceeding	-

4. TITLE AND SUBTITLE

Classification of non-time-locked rapid serial visual presentation events for brain-computer interaction using deep learning

5a. CONTRACT NUMBER
W911NF-14-1-0043

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
206022

6. AUTHORS

Vernon Lawhern, Lenis Mauricio Merino, Kenneth Ball, Li Deng, Brent J. Lance, Kay Robbins, Yufei Huang, Zijing Mao

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

University of Texas at San Antonio
One UTSA Circle

San Antonio, TX 78249 -1644

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
64697-LS-REP.2

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

Deep learning solutions based on deep neural networks (DNN) and deep stack networks (DSN) were investigated for classifying target images in a non-time-locked rapid serial visual presentation (RSVP) image target identification task using EEG. Several feature extraction methods associated with this task were implemented and tested for deep learning, where a sliding window method using the trained classifier was used to predict the occurrence of target events in a non-time-locked fashion.. The deep learning algorithms explored based on deep stacking networks were able to improve the error rate by about 5% over existing algorithms such as linear

15. SUBJECT TERMS

brain-computer interfaces, electroencephalography image, classification, learning (artificial intelligence), neural nets

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Yufei Huang
UU	UU	UU	UU		19b. TELEPHONE NUMBER 210-458-6270

Report Title

Classification of non-time-locked rapid serial visual presentation events for brain-computer interaction using deep learning

ABSTRACT

Deep learning solutions based on deep neural networks (DNN) and deep stack networks (DSN) were investigated for classifying target images in a non-time-locked rapid serial visual presentation (RSVP) image target identification task using EEG. Several feature extraction methods associated with this task were implemented and tested for deep learning, where a sliding window method using the trained classifier was used to predict the occurrence of target events in a non-time-locked fashion.. The deep learning algorithms explored based on deep stacking networks were able to improve the error rate by about 5% over existing algorithms such as linear discriminant analysis (LDA) for this task. Initial test results also showed that this method based on deep stacking networks for non-time-locked classification can produce an error rate close to that achieved for time-locked classification, thus illustrating the power of deep learning for complex feature spaces.

Conference Name: 2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP)
Conference Date: July 08, 2014

CLASSIFICATION OF NON-TIME-LOCKED RAPID SERIAL VISUAL PRESENTATION EVENTS FOR BRAIN-COMPUTER INTERACTION USING DEEP LEARNING

Zijing Mao[1], Vernon Lawhern[2,3,5], Lenis Mauricio Meriño[1], Kenneth Ball[2,5], Li Deng[4], Brent J. Lance[5], Kay Robbins[2], Yufei Huang[1]

1. Department of Electrical and Computer Engineering, 2. Department of Computer Science,
University of Texas at San Antonio, USA
3. DCS Corporation, USA, 4. Microsoft Research, USA
5 Translational Neuroscience Branch, U.S. Army Research Laboratory, USA

ABSTRACT

Deep learning solutions based on deep neural networks (DNN) and deep stack networks (DSN) were investigated for classifying target images in a non-time-locked rapid serial visual presentation (RSVP) image target identification task using EEG. Several feature extraction methods associated with this task were implemented and tested for deep learning, where a sliding window method using the trained classifier was used to predict the occurrence of target events in a non-time-locked fashion.. The deep learning algorithms explored based on deep stacking networks were able to improve the error rate by about 5% over existing algorithms such as linear discriminant analysis (LDA) for this task. Initial test results also showed that this method based on deep stacking networks for non-time-locked classification can produce an error rate close to that achieved for time-locked classification, thus illustrating the power of deep learning for complex feature spaces.

Index terms - RSVP, non-time-locked events, feature selection, deep learning, deep neural networks, deep stacking networks, brain-computer interaction.

I. INTRODUCTION

A brain computer interaction (BCI) system allows human subjects to communicate with or control an external device with their brain signals [1], or to use those brain signals to interact with computers, environments, or even other humans [2]. One application of BCI is to use brain signals to distinguish target images within a large collection of non-target images [2]. Such BCI-based systems can drastically increase the speed of target identification in large image databases over manual procedures [3]. Data collection for training such BCI systems is commonly carried out using the rapid serial visual presentation (RSVP) paradigm [2], where test subjects are asked to identify a target image from a continuous burst of image clips presented at a high rate. The EEG recordings are collected and a classifier capable of predicting the presence of target images based on EEG responses is trained using this data.

Classification for RSVP data is usually performed in a 'time-locked' fashion, by analyzing the spectrum or amplitude in the EEG signal 300-1000ms immediately following presentation of target and non-target images. Although processing RSVP data without time locking is more realistic, non-time-locked classification is significantly more demanding because target event timing needs to be explicitly or implicitly estimated. So far, a number of classifiers including logistic regression, linear discriminant analysis (LDA) and support vector machines (SVM) have been proposed in the literature to address the classification of time-locked events [2, 4]. These classifiers have been reported to provide less than 10% error rate. The problem of imperfectly time-locked events was considered in [5], where event timing was assumed to be unknown but occurring within a small known interval. Performance close to that achieved under perfect time locking were reported. However, classification for completely non-time-locked events has as of yet not been addressed.

In this paper, we investigate deep learning (DL) solutions to non-time-locked RSVP classification. Deep learning is a term for a new family of learning methods that have been shown to offer superb representation of complex data by using a multiple-layered architecture [6, 7]. DL has gained great interest in recent years due to its ability to outperform alternative classification methods in several machine learning competitions and in a variety of applications, including image classification and speech recognition [8, 9]. However, DL applications for EEG data analysis are at a very early stage. It is not clear if and how unique characteristics of EEG data including high dimensional feature spaces, temporal and spatial data correlation, and excessive noise will affect the implementation and performance of DL algorithms. The goal of this work is two-fold. First, we aim to develop solutions to classify non-time-locked events in RSVP. Second, we intend to investigate the use of DL algorithms for EEG data analysis in BCI research.

II. MATERIAL AND METHODS

A. Experimental Design

The RSVP EEG recordings were obtained from [2], which include brain activities of five participants presented with a series of bursts of images in an RSVP paradigm. Each burst lasts for 4.1s and consists of 49 images presented at a speed of 12 images/second. A burst may contain zero or one target images, where a target image includes a silhouette of airplane which is not present in non-target images. To ensure no interference from burst edges, the target image is not presented within 500 milliseconds (ms) from the onset and offset of the burst. EEG recordings were collected using a BIOSEMI ActiveTwo system with 256 electrodes at 256 Hz sampling rate with 24-bit digitization.

B. Data Preprocessing and Prediction Objective

The raw data include 7.1s EEG data epochs, each centered on a 4.1s RSVP burst. The data was first bandpass-filtered in the range 2-100Hz. Independent component analysis (ICA) using the Extended Infomax Algorithm in EEGLAB [10] was then performed to reduce the correlation between channels and to remove noise. The 16 components with highest variance were retained. To capture the time-frequency space characteristics, a wavelet transformation was applied to the ICA transformed data to obtain a temporal-IC power spectrum for 18 frequency bands evenly sampled on a logarithmic scale from 2-100Hz. Only 5s of the 7.1s data epochs were extracted from the recording of subject 1-5, where for target epochs, the target event onset time was at 2s of the 5s epoch. There were a total of 138, 129, 114, 121, and 145 target epochs for subject 1 to 5, and 188, 171, 194, 188, and 190 non-target epochs for subject 1 to 5, respectively. Our goal is to predict if and when a target image is present in a 5-second epoch. The transformed data of each epoch represented the power of EEG recording distributed in three dimensions: independent components (ICs), frequency, and time.

C. Construction of training data.

We developed a solution based on sliding windows, where a 500-ms long window slides from the beginning to the end of an epoch at a step size of 1 sample. For each slide location, the EEG data within the window is subjected to a classifier to predict if a target image is present. The key to this solution is to train a classifier that can predict a target event if the event-related brain response occurs within the 500ms window of the input EEG data. What makes this training difficult is that the response can happen at any place within the window. To construct a training set, we define the target event region as the region from the event onset time (2s) to one second (1s) afterwards. Since a target image can appear 200-300ms after the onset time and a target-related potential is known to occur 300-500ms after an image appears in each of the target epochs, a 500ms window within the target event region should contain event-related brain response. To account for potential offset between the sliding window and the target event, we also investigated the target event region with 200ms offset, which covers 200ms before target onset to 300ms after target onset. For each of the defined target event regions, 50 random sections of 500ms-long EEG data were randomly taken from the one-second target event region and labeled as target event. Next, 50 non-target labeled data samples of 500ms windows were extracted randomly from the 5-second non-target epochs. This was done for each target/non-target epochs and in the end, two training datasets corresponding to offsets of 0s and 200ms, respectively, were obtained for each subject. Each set includes 6750, 6450, 5700, 6050, 7250 target event samples and 9250, 8550, 9700, 9400, 9500 non-target event samples for subject 1 to 5, respectively, where a sample contains IC-time-frequency powers for 500ms resulting in a 36864 (16 ICs ×18 frequencies × 128 time samples) dimensional feature vector. The classification output includes two labels: target or non-target event.

D. Deep Learning Algorithms

Deep learning is a term that refers to a class of new machine learning algorithms that exploit architectures of layered modules of supervised or unsupervised learning algorithms. Depending on the learning nature of each module, the existing deep learning algorithms can be classified into generative, discriminative, and hybrid architectures [7]. Generative architectures include the deep auto-encoder, which consists of layers of neural networks, whose output has the same dimension as the input. The main objective of deep auto-encoders is to extract features from data as opposed to classification. Deep stacking networks (DSN) exemplify discriminative architectures [11-13]. For DSN, each module is a classifier that takes a simplified multilayer perceptron, which includes a shallow sigmoidal neural network followed by a linear classifier. The hybrid class includes the well-known deep neural network (DNN), which consists of layers of restricted Boltzmann machines with a classification module at the very top. A survey of these algorithms can be found in [6].

In this paper we will focus on the DNN and the DSN. Two variations of DSN, one with a linear (DSN-L) and one with a sigmoidal activation function (DSN-S) were implemented. When applying DL to EEG data, high feature dimension and potentially strong temporal, spatial, and frequency feature correlations can be problematic. The dimension of the input features in the described training set is 36864 (16×18×128), which is unrealistically high for most classifiers, including DL algorithms. Reduction of feature dimension needs to be performed before these features can be used for DL-based classification. Moreover, the features in adjacent channels, frequency bands, and time samples are highly correlated, which might affect the convergence of DL algorithms. We used ICA to reduce the spatial (EEG channels) feature dimension and correlation. We reduced the dimension and correlation of temporal and frequency features using down-sampling and principal component analysis.

III. RESULTS

Feature dimension reduction schemes including down-sampling and principal component analysis (PCA). Parameters for preprocessing and DL algorithms were identified using grid search on the data from subject 1. Cross-validated training and prediction of non-time-locked target events were conducted for each of the remaining 4 subjects using the dimension reduction scheme and DL parameters that led to the best performance in subject 1.

A. Investigation of feature dimension reduction

1) Reduction by down-sampling

The impact of down sampling on classification was tested by comparing results from 4 different orders of reduction (i.e., down sampling by a factor 2, 4, 8 or 16). LDA was used as the baseline classifier for this investigation. Table I shows the average error rates for 5-fold cross validation using different down-sampling factors. The lowest error rate was obtained for a down-sampling factor of 8. As a result, the time samples were reduced by a factor of 8 from the original 128 to 16, reducing the feature dimension to 4608.

Down sampling factor	2	4	8	16
Error rate	0.3724	0.3657	0.2959	0.2981

Table I. Error rate of LDA for different downsampling factors

2) Reduction by principal component analysis

Both DSN and DNN were tested on the down-sampled data, however neither resulted in convergence. Correlation of features from adjacent frequency bands and time points contributed to the poor convergence of the stochastic optimization used in these DL algorithms. To overcome this problem, principal component analysis was applied to reduce the correlations among features over time and frequency, further reducing feature dimension. Table II demonstrates the classification performance of LDA both before PCA and after using only the first 100 principal components (PCs). It is clear that applying PCA improves the classification performance of LDA. We use the first 100 PCs based on the results of grid search.

Dimension reduction	Before PCA	After PCA (100) PCs)
Classifier	LDA	LDA
Error rate	0.2959	0.2044

Table II. Error rate before and after PCA

B. Investigation of deep learning parameters

For both DNN and DSN, the number of layers and the number of hidden units in each layer can impact the classification performance. Tuning parameters including the learning rate are also important factors. We determined the best number of hidden units and layers as well as the values of other tuning parameters by using the training data of subject 1 with 5-fold cross validation.

DNN	First (100 units) layer	First (100 units) + second (50 units) layer
Error rate	0.1791	0.1706

Table III. Error rate of DNN with different hidden layers

For DNN, two-layer architectures resulted in the best performance (Table III); using more than 3 layers resulted in poor convergence. For DSN, the best performance was achieved at 21 layers with 60 hidden units for 0ms offset and 14 layers with 60 hidden units for 200ms offset. The best performance for different DL algorithms was also summarized in Table IV. For both DNN and DSN, only one iteration of fine tuning was applied. The performance of DNN and DSN was similar, with DSN-S achieving a slightly lower error rate. The results for 0s offset are also consistently better than those for 200ms offset, suggesting that training data with 0s offset better captures the brain response to target image. It is likely that some of the target event training samples for 200ms offsets contained no corresponding brain response and thus were mislabeled. The resulting values for DL algorithms are fixed for training and prediction in the remaining 4 subjects in later sections of the paper.

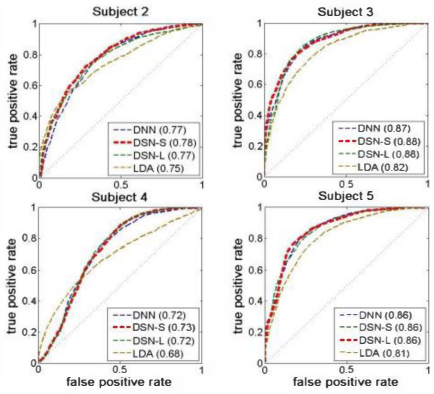

Figure 1. ROC curves for subject 2-5

Offset	DSN-L	DSN-S	DNN
0ms	0.1693	0.1666	0.1706
200ms	0.1883	0.1856	0.1907

Table IV. Average error rates of DL algorithms

C. Training of DL classifiers for individual subjects

We proceeded to train separate DL classifiers for subject 2 to 5. The parameter settings obtained in Section III.B that led to the best performance were selected for the training. The goal of training was to estimate the DL weights and compare the DL performance with LDA. 5-fold cross validation used for evaluation.

Figure 1 shows the ROC curves and Area-Under-the-Curve (AUC) statistics of the trained DL algorithms for all 4 subjects for 0s offset. Consistent with results reported previously [5, 6], the AUC performance for subjects 2 and 4 was lower than that for subjects 3 and 5. The three DL algorithms achieved similar performance. All the DL algorithms improved performance by about 5% in AUC for all subjects when compared with LDA.

D. Predictions of non-time-locked target events

We proceeded to predict the target event in a 5-second epoch with sliding windows using the trained DSN-S classifiers for each subject (2-5), respectively. A data window in the target event region of a target 5s-epoch was labeled as a "target event" while those in the remaining regions were labeled "non-target events". The prediction ROC curve is shown in Fig. 2-A and the AUC statistics are very close to those of the training in Fig. 1, suggesting that our constructed training dataset was sufficient to capture the characteristics of the EEG data for both target and non-target events. Once again, the performance for subjects 3 and 5 is better. This performance is also comparable to that of the time-locked prediction [4]. Examples of the prediction results in both target and non-target epochs are shown in Fig. 3. We observed that our method did very well in predicting the target event regions and could also correctly predict the onset of the target event. Overall, the false negative predictions were small and made towards the end of the target

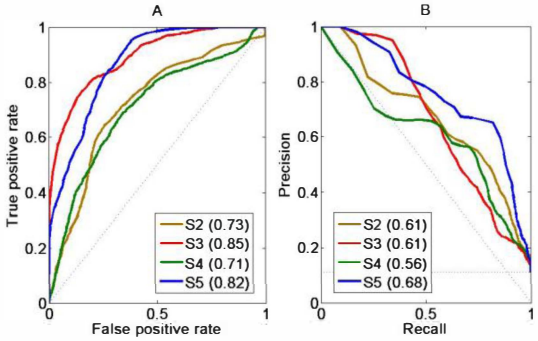

Figure 2. ROC (A) and Precision-Recall (B) curves of prediction by sliding window.

event region. It is likely that towards the end of target event region, the brain response to target images has already faded, resulting in the false positive predictions.

To further investigate the effectiveness of our method in the prediction of target events, we plotted in Fig. 2-B the precision-recall (PR) curve for the predictions made in the target event epochs. The precision is defined as the percentage of correctly predicted target events among all data windows that were predicted as a target event at a given decision threshold. As can be seen, except for subject 4, DSN-S can achieve and maintain 100% precision until the prediction recall reaches almost 10%. This implies that top 10% highly ranked predictions are true positive prediction of target event. Taken together, these preliminary results indicate that the proposed sliding window method for predicting non-time-locked target events may be able to achieve a performance level close to that of the time-locked prediction (Table V).

Subject	2	3	4	5
DSN	0.73	0.85	0.71	0.82
cLDA	0.81	0.91	0.68	0.88

Table V. Comparison of AUCs between the proposed DSN-based sliding window method for non-time-locked event prediction (Fig. 2) and cLDA for time-lock event prediction reported in [5].

IV. Conclusion and Future Work

We presented in this paper an investigation of deep learning classifiers based on the architectures of the DSN and DNN for automatic classification of non-time-locked image RSVP events. The preliminary results obtained from analyzing five subjects, one for training and four for validation, indicate that deep learning may be able to improve the prediction error rate by about 5% over other existing mainstream methods for this task. In addition, we provided preliminary results showing that a sliding window method based on the DSN produced an error rate similar to that for prediction of time-locked events. Our study in this paper suggests that deep learning has a strong potential to be a powerful tool for BCI research. However,

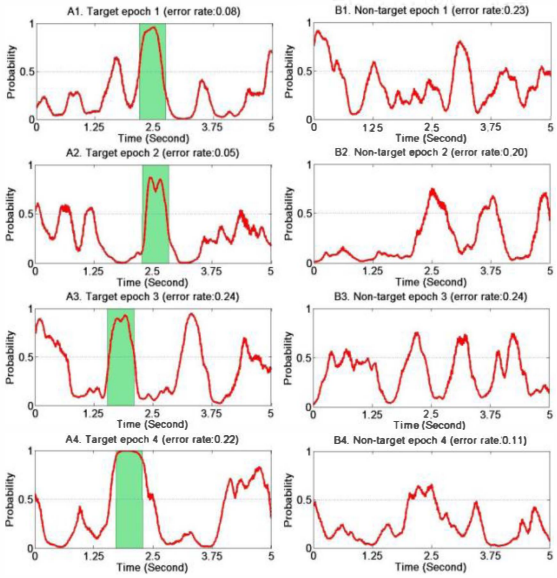

Figure 3. Examples of predictions by sliding-window. The left column includes epochs with target events. The target event regions are highlighted by green. The vertical axis denotes the probably of predicting a target event. The line at 0.5 represents the decision threshold

careful extraction of features from the EEG signal to feed into the existing DSN or DNN architectures are likely to further improve the classification accuracy in this application domain. How to incorporate the stage of modeling EEG features into the deep learning architecture will be a topic of future study. The main challenge is to take into account both temporal and spatial correlations in the observed data exhibiting variable dimensionality. This is a popular topic in deep learning research with other application domains [17, 18, 19] that is likely to help the BCI application discussed in this paper.

V. Acknowledgements

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0022. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implies, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for the Government purposes notwithstanding any copyright notation herein. This work received computational support from Computational System Biology Core at the University of Texas at San Antonio, funded by the National Institute on Minority Health and Health Disparities (G12MD007591) from the National Institutes of Health. We also thank Dr. Scott Makeig and Dr. Nima Bigdely-Shamlo from the University of California at San Diego for sharing with us the RSVP data.

VI. Reference

1. Wolpaw, J.R., et al., *Brain-computer interface technology: a review of the first international meeting.* IEEE Trans Rehabil Eng, 2000. **8**(2): p. 164-73.
2. Bigdely-Shamlo, N., et al., *Brain activity-based image classification from rapid serial visual presentation.* IEEE Trans Neural Syst Rehabil Eng, 2008. **16**(5): p. 432-41.
3. Sajda, P., et al., *Cortically-Coupled Computer Vision*, in *Brain-Computer Interfaces*. 2010, Springer. p. 133-148.
4. Meng, J., et al., *Characterization and robust classification of EEG signal from image RSVP events with independent time-frequency features.* PLoS One, 2012. **7**(9): p. e44464.
5. Meng, J., et al., *Classification of Imperfectly Time-Locked Image RSVP Events with EEG Device.* Neuroinformatics, 2013.
6. Deng, l. and D. Yu, *Deep Learning for Signal and Information Processing*, in *Foundations and Trends in Signal Processing*. 2013, NOW Publishers.
7. Hinton, G.E., S. Osindero, and Y.W. Teh, *A fast learning algorithm for deep belief nets.* Neural Computation, 2006. **18**(7): p. 1527-1554.
8. Hinton, G., et al., *Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups.* Signal Processing Magazine, IEEE, 2012. **29**(6): p. 82-97.
9. Bengio, Y., *Learning deep architectures for AI.* Foundations and trends® in Machine Learning, 2009. **2**(1): p. 1-127.
10. Delorme, A. and S. Makeig, *EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis.* Journal of neuroscience methods, 2004. **134**(1): p. 9-21.
11. Deng, L., D. Yu, and J. Platt. *Scalable stacking and learning for building deep architectures*. in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. 2012. IEEE.
12. Hutchinson, B., L. Deng, and D. Yu, *Tensor deep stacking networks.* IEEE Trans Pattern Anal Mach Intell, 2013. **35**(8): p. 1944-57.
13. Deng, L. and D. Yu. *Deep convex net: A scalable architecture for speech pattern classification*. in *Proceedings of the Interspeech*. 2011.
14. Dahl, G., et al. "Context-dependent, pre-trained deep neural networks for large vocabulary speech recognition," IEEE Trans. Audio, Speech, & Language Proc., Vol. 20, pp. 30-42, January 2012.
15. Krizhevsky, A., Sutskever, I. and Hinton, G. "ImageNet classification with deep convolutional neural Networks," Proc. NIPS 2012.
16. Yu, D., et al. "The Deep Tensor Neural Network with Applications to Large Vocabulary Speech Recognition," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, pp. 388-396, 2013.
17. Deng, L. and Chen, J. "Sequence Classification Using the High-Level Features Extracted from Deep Neural Networks," Proc. ICASSP, 2014.
19. Chen, J. and Deng, L. "A Primal-Dual Method for Training Recurrent Neural Networks Constrained by the Echo-State Property", arXiv:1311.6091, pp. 1-16, 2013.
20. Vinyals, O., et al. "Learning with recursive perceptual representations," Proc. NIPS, 2012